

電子情報の長期的保存とアクセス手段の確保

澤田 大祐

国立国会図書館関西館事業部電子図書館課

〒619-0287 京都府相楽郡精華町精華台 8-1-3

Tel: 0774-98-1476, Fax: 0774-94-9118, E-Mail: d-sawada@ndl.go.jp

概要

電子情報の保存と再生は技術的環境に依存するものであり、紙上の情報に比べて脆弱である。国立国会図書館では平成14年度より電子情報の長期保存に関連した調査を実施している。平成16年度の調査では、平成11年以前より所蔵するCD-ROMからサンプルを抽出してマイグレーションとエミュレーションを行い、最新の環境下での再生を試みた。その結果、ほぼ全てのサンプルについてマイグレーションに成功したが、エミュレーションに成功したのはサンプルの3割に留まった。

キーワード

電子情報の保存、電子情報の脆弱性、パッケージ系電子出版物、マイグレーション、エミュレーション

Ensuring Long-term Preservation and Access of Digital Information

Daisuke SAWADA

Digital Library Division, Projects Department, Kansai-kan of the National Diet Library

8-1-3 Seikadai, Seika-cho, Soraku-gun, Kyoto, 619-0287, JAPAN

Tel: +81-774-98-1476, Fax: +81-774-94-9118, E-mail: d-sawada@ndl.go.jp

Abstract

Preserving digital information and rendering it depends entirely on the technical environment. Thus digital information is said to be very fragile compared to paper-based materials. The National Diet Library began a survey of digital preservation from FY2002. In FY2004, the survey was made on the CD-ROMs sampled at random that had been received up to FY1999 by the NDL to see how they could be worked for migration or emulation. As a result, it was found that migration could be done without significant problems. However, emulation was valid for only about 30% of all samples.

Keywords

preservation of digital information, fragility of digital information, packaged digital publications, migration, emulation

1. はじめに

1990年代中頃より、欧米やオセアニア各国の国立図書館を中心として、電子情報の長期的保存に関する研究が行われるようになった。ポーンデジタル資料の増加と質的多様化により、紙資料だけではなく電子情報についても長期的な保存とアクセスを保証する必要性が生じたのである。その後研究が進むにつれ、電子情報の保存には、紙資料の保存とは異なる様々な問題点があることが明らかになった。

2003年10月、国際連合教育科学文化機関（UNESCO: United Nations Educational, Scientific and Cultural Organization）は、「デジタル遺産の保護に関する憲章」を採択した[1]。「憲章」では、インターネット情報を含め、現代社会において重要な電子情報の保存が保証されていない状況にかんがみ、各国政府において、問題意識の喚起と保存のための取り組みが必要であることを宣言している。

国立国会図書館（National Diet Library、以下NDLとする）においても平成14年度より調査研究を開始した。NDLでは、平成12年10月1日の国立国会図書館法の一部改正法施行に伴い、パッケージ系電子出版物を納本対象として受け入れるようになった[2]が、その以前から購入などによって収集した資料も所蔵している。NDLが収集し後世に遺す出版物は紙媒体に限定されるものではなく、電子出版物も貴重な文化遺産である。また、NDLでは明治時代以前の図書の電子化[3,4]やネットワーク系電子出版物の収集[5]など、電子情報の収集、保存、提供に関する様々な事業を進めており、所蔵する電子情報は今後急増していくことが予想される。電子情報の長期的保存と将来にわたる情報アクセスの保証は重要な問題である。

本稿では、平成16年度の調査結果を中心として報告する[6]。

2. 長期的保存の問題点

「電子情報は利用によって劣化するものではないので、半永久的に保存できて便利だ」「電子情報はコピーが簡単にできるので、長期保存は容易だ」という考え方がある。しかし実際には、紙に印刷された情報に比べて脆弱であり、長期保存とアクセス手段の確保という観点からは様々な問題がある。NDLが平成15年度に行った調査では、平成11年度までに受け入れた電子資料（注）のサンプルのうち約7割について、現在の最新環境下で利用するとき何らかの問題が発生することが明らかになった。（表1）

利用が不可能となる原因は、次のように大別される。

(1) OSの不適合

調査において、資料が利用できなかった最大の理由は、OS（Operating System、基本ソフトウェア）に係るものである。異常終了、一部機能の不作動、インストールの失敗などの問題が発生した。電子資料を再生する場合、機器と再生用のアプリケーションソフトウェア（その電子資料自体が再生対象のアプリケーションソフトウェアであることも多い）を必要とするが、通常、アプリケーションソフトウェアは特定のOSでのみ動作が保証されている。新しい種類やバージョンのOSが登場すれば、新しいアプリケーションソフトウェアも続々と登場する。その一方で、市場での競争力を失い旧式化していくOSも存在する。OSの不適合が原因で利用不可となる電子資料は、新規OSの増加に伴って今後増加することが予想される。

(2) アプリケーションプログラムに関する問題

電子書籍用のプログラムや、古いバージョンのアプリケーションプログラムが必要であるため、標準的なパーソナルコンピュータの環境では利用できない資料も多くあった。この場合、必要とするアプリケーションプログラムをインストールすれば利用が可能であると予想されるが、旧式化したアプリケーションプログラムを入手することは困難である。永続的な文書保存に適しているとされるファイル形式もあるが、対応するアプリケーションプログラムが供給され続ける保証はない。

表1 電子資料の受入時期ごと利用可能性

受入年度	調査資料点数	利用可能資料点数	利用可能資料の割合
平成2年度以前 (-1991.3)	29	1	3%
平成3年度 (1991.4-1992.3)	20	2	10%
平成4年度 (1992.4-1993.3)	20	2	10%
平成5年度 (1993.4-1994.3)	20	3	15%
平成6年度 (1994.4-1995.3)	20	3	15%
平成7年度 (1995.4-1996.3)	20	8	40%
平成8年度 (1996.4-1997.3)	20	8	40%
平成9年度 (1997.4-1998.3)	20	16	80%
平成10年度 (1998.4-1999.3)	20	13	65%
平成11年度 (1999.4-2000.3)	11	6	55%
合計	200	62	31%

(3) 媒体とハードウェアの旧式化

記録媒体関連の問題により利用できない電子資料もあった。その大半が5インチフロッピーディスクであり、対応ドライブ機器を用意できないために読み取りができなかった。ドライブ機器が対応しているはずの3.5インチフロッピーディスクとCD-ROMについても読み取れないものがあった。

電子媒体に新たな規格が出現して普及すると、旧式の電子媒体の利用が急激に困難になる。例えば、8インチ及び5インチのフロッピーディスクドライブ、生産を中止したゲーム機用の媒体などは、読み出し装置の入手が難しいため、利用は困難である。仮に読み出し装置を入手できたとしても、媒体・装置のどちらかが耐久年限を過ぎてしまえば、利用できない。

さらに、媒体そのものの劣化も問題である。特に、近年では光ディスクの耐久性について注目が高まっている。開発当初は「半永久的な保存メディア」とされていたCDであっても、数十年程度で使用が不可能なほど劣化するという調査結果もある。[7]

3. マイグレーションとエミュレーション

電子情報の長期的保存に関する様々な問題を解決する手段として、マイグレーションとエミュレーションが考えられる。

3.1 マイグレーション

マイグレーションとは、プログラムやデータの移行および変換作業であって、同種の新しい媒体に移し替えること、異種の媒体に移し替えること、記録形式を最新のものにすること、アプリケーション・ソフトウェアを他の環境で動作させるために作り直すことを意味する。

マイグレーションには様々な手法が含まれるが、ここでは次のように分類する。

- (1) 同種の新しい媒体に移し替えること
- (2) 異種の媒体に移し替えること
- (3) データ形式を変換すること
- (4) プログラムやシステムを新規環境に移行すること

(1)(2) は、電子情報自体を長期的に保存するための行為、(3)(4) は、電子情報へのアクセス手段を確保するための行為といえることができる。

(1)(2) は作業としては容易といえるが、所蔵資料の規模に応じて作業量が増大する。また、個々の電子出版物を大規模記憶装置に移行して集約しない限り、非常にコストのかかる作業になる。しかも、媒体に記録されている情報に確実にアクセスできる段階で次の媒体に移さなくてはならないため、媒体寿命といわれる年数より短い間隔で作業を繰り返さざるを得ない。(3) は、JPEG を JPEG2000 に変換する、RTF を PDF に変換するなどのデータ形式の変換作業である。変換先データ形式が標準的なものであって、今後も広く使われつづけるとされる形式であれば、再生は確保できる。しかし、原本性が損なわれるという問題が伴う。(4) は、仕様書や設計書、ソースプログラム一式を揃え、必要個所の修正を行い、プログラムを再作成することである。商用ソフトウェアの必要物一式の入手は困難なので、適用対象は組織内部で作成したプログラム等に限定されると思われる。マイグレーションというカテゴリーに含めているものの、他の方法と比べ、作業内容も、必要とされる技術レベルも大きく異なる。

3.2 エミュレーション

エミュレーションとは、エミュレータと呼ばれるソフトウェアを用い、コンピュータが持つ OS 環境上で別の OS やハードウェアが持つ環境を実現させることである。例えば、Windows XP が作動しているパーソナルコンピュータにエミュレータを搭載することで、Windows XP 上に新たな「仮想パソコン」を作り、その上で Unix や旧式の Windows など別の OS を起動させることができる。(図 1)

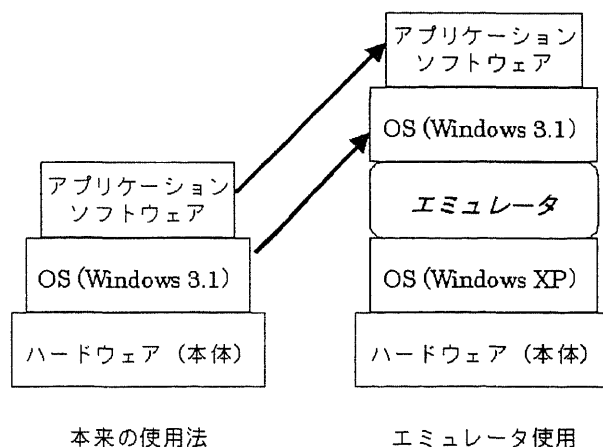


図 1 エミュレータの役割

電子資料を再生するためには、そのための環境、すなわち特定のハードウェアとソフトウェアが必要である。しかし、ハードウェアの寿命は短く、動作可能な状態で保持し続けていくことは現実的でない。エミュレータはアプリケーションソフトウェアの位置にあって、擬似的に、旧式ソフトウェアが必要としている

機能を再現する。ハードウェアや OS が移り変わっても、エミュレータのみ作成すればそれまでのアプリケーションソフトウェアや OS を利用することができる。このようなエミュレータは多数作成され、頒布されている。しかし、独自に作成するためには高度な技術が必要であり、また完全に旧式環境を再現できるとは限らない。

4. 平成 16 年度の調査

2章で述べた様々な問題の解決に、電子情報の長期的保存におけるマイグレーション、エミュレーション及びファイルのデータ形式変換が有効であるかどうか、調査を行った。調査の手順は、将来 NDL で業務を行う上で妥当な実施形態を想定し、図 2 に示す手順で行った。

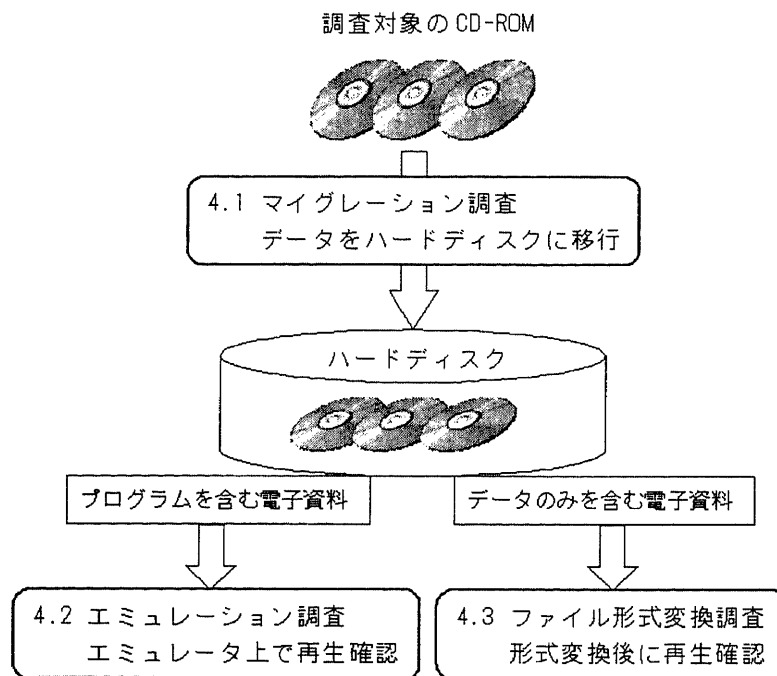


図 2 調査の進め方

4.1 マイグレーション調査

Windows 用または MS-DOS 用のソフトウェアを収録した CD-ROM/R/RW で、NDL が平成 11 年度以前に受け入れた国内刊行の電子資料の中から、マイグレーション用のサンプル 354 点を選択した。各サンプルについて、市販のマイグレーションプログラムを用い、CD からハードディスクへのマイグレーションを行った。その結果、354 点のうち 352 点についてマイグレーションが成功した。残りの 2 点は、処理開始から 1 時間経過してもマイグレーションを終了することができなかった。この原因について調査した結果、コピープロテクトが施されていたことが判明した。

4.2 エミュレーション調査

マイグレーションに成功した 352 点の中から、プログラムを含む資料を 100 点選択し、エミュレーション調査用のサンプルとした。また、OS に Windows XP Professional を搭載したパーソナルコンピュータに市販のエミュレータをインストールし、さらに各種の旧式 Windows OS をインストールして、電子資料の再生に必要な旧式の OS 環境を構築した。各サンプルを旧式 OS 環境上で再生し、表 2 に示す機能について動作確認を行った。

表 2 エミュレーションの再生確認内容

No.	機能種類	機能の有無調査	再生確認内容
1	インストール	当該資料の再生にはインストールが必要か否か	インストール処理が正常に終了するか
2	起動画面	起動直後に表示される画面の有無	正常に起動して起動画面が表示されるか
3	指示に対する動作	各ボタン、エンターキーなどのオペレーションの有無	左記の各動作に対して正しく作動するか
4	静止画像	写真、テキスト、スライドなどの静止画像の有無	文字化け、画像の乱れなどの有無
5	動画	アニメーション、ムービーなどの動画の有無	文字化け、画像の乱れなどの有無
6	音声	音声の有無	雑音などの有無
7	計算、判定	計算、判定などのロジック有無	正常に計算されるかなど
8	検索抽出	検索、抽出機能の有無	検索、抽出結果がユーザの指示どおりか
9	データ記録	次回起動時までのデータ記録機能の有無	データが記録されるか
10	その他	上記以外の機能があれば、その機能概要を調査	左記の機能が正しく動作するか

その結果、全体の 7 割の資料で、再生時に何らかの問題が発生した。図 3 は、電子資料が必要とする OS ごとの結果である。

(1) Windows 98

エミュレータ上に Windows 98 をインストールし、12 点の電子資料について動作確認を行った。今回の調査の中では比較的新しい OS であり、音声の再生を正しく行えなかった媒体が 1 点あったが、残りは再生に問題がなかった。

(2) Windows 95

エミュレータ上に Windows 95 をインストールし、55 点の電子資料について動作確認を行った。このうち再生が可能なものは 16 点に留まった。一部の再生が不可能なものについては、ほとんどが音声の問題（雑音、再生速度の乱れ）であった。

(3) Windows 3.1 及び DOS

エミュレータ上に DOS(J6.1/V) と Windows 3.1 をインストールし、Windows 3.1 用電子資料 31 点、DOS 用電子資料 2 点について動作確認を行った。Windows 3.1 上で必要とされる音声用ドライバが入手できなかったこと、画面サイズの不適合などの原因により、再生が可能なものは 2 点のみであった。

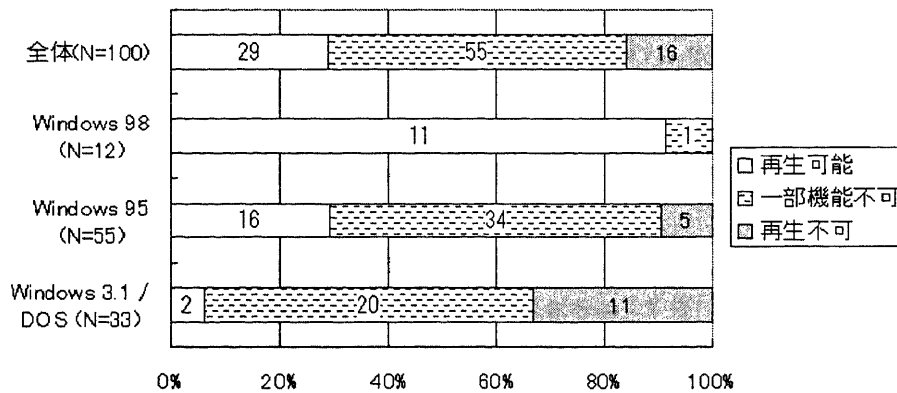


図3 エミュレーションによる再生の結果

4.3 ファイル形式変換調査

データのみを含み、プログラムを含まない電子資料については、特定の OS ではなく、特定のアプリケーションソフトウェアに再生環境を依存しているものが多い。この問題を回避するためには、別のアプリケーションソフトウェアでも再生できるように、あらかじめデータ形式を変換しておくことが有効であると考えられる。

4.1 においてマイグレーションを行った 352 点の中から、データのみを含む資料を 100 点抽出し、ファイル形式変換調査用のサンプルとした。サンプルの内容は、次のように大別される。(図 4)

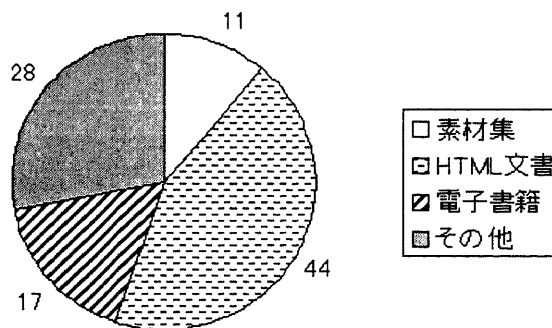


図4 データのみを含む資料の分類

(1) 素材集など

多種多様なファイル形式を含むが、それぞれのファイルに関連性がなく、個別に再生するもの。電子資料に含まれる多数のファイルから、利用したい 1 ファイルを取り出して使用する。

(2) ブラウザで再生する HTML 文書

HTML 形式のファイルは、様々な形式のデータファイルが関連付けされて再生される。複数の HTML 形式ファイルの組み合わせによる関連付けの他にも、静止画・動画・音声・その他アプリケーションのデータが文書内に埋め込まれている場合が多い。

(3) 電子書籍・辞書

電子書籍類は、データ本体とその制御情報によって構成されている。付随する制御情報によって、データの内容に合わせた表現が行われたり、データの検索を行えるようになっている。例えば、文字情報に音声情報を連携させることによって、辞書に描かれた動物の挿絵をクリックすると鳴き声が流れる、といった表現が実現されている。EPWING フォーマット（JIS X4081「日本語電子出版検索データ構造」）などがこれに該当する。

(4) その他

上記いずれにも当てはまらないもの、例えば、PDF 形式のファイルに動画のファイルが付随しているものなどがあつた。

一般的なファイル形式変換ソフトウェアでは、変換が1ファイルごとに行われる。そのため、(2)に該当するファイルについて変換を行うと、連携している他のファイルとの関連性が失われてしまうために、元のコンテンツを一体的に再生することができなくなる。また、(3)に該当するファイルも、制御情報を用いてデータが連携されているために、個別にファイルを取り出して利用することができない。そのため、(2)や(3)についてファイル形式変換を行うことは、現状では困難である。

本調査では、(1)・(4)に該当する電子資料39点のうち、ISOによって規格化されているものなど、ソフトウェアベンダーに依存しない形式のファイルのみによって構成されている資料が24点あつた。これらは長期的に利用保証することが比較的容易であると考え、残りの15点について、市販のファイル形式変換ソフトウェアを用い、より規格化された形式に変換することを試みた。変換作業の例を表3に示す。

表3 ファイル変換作業の例

	資料に含まれるファイルの形式		ファイル形式 変換調査	結果
	規格化されている	されていない		
サンプル1	HTML, BMO, ICO, WAV	なし	対象外	
サンプル2	TXT	XLS	対象	変換成功
サンプル3	BMP, JPG	AI, CXT, DXR	対象	変換失敗

その結果、ファイル形式の変換に成功したのは2点のみであつた。内訳は、DOC（Microsoft Word 文書）形式からPDF（Adobe）形式、XLS（Microsoft Excel 文書）形式からPDF形式が各1点である。

5. 考察

5.1 マイグレーションによる電子情報の長期的保存

4.1のマイグレーション調査では、ほとんどの資料について異種媒体への情報移行が可能であることが明らかになった。情報記憶媒体の進歩に合わせて内容を異種媒体に移し変えることは、最も容易かつ確実な保存手段である。しかし、これと同時に、NDLが所蔵する電子資料の中には、コピープロテクトのためにマイグレーションできない資料があることも明らかになった。このような資料を放置すれば、媒体の劣化がそのまま内容の消滅に結びつく危険性が高い。また、電子媒体に関する技術の進歩に伴い、コピープロテクトの技術も日々変化している。さらに近年では、デジタル著作権管理（DRM: Digital Rights Management）についても様々な技術が開発されており、ネットワークを介して得られるコンテンツについてもライセンス

ス管理が長期的資料保存の妨げになる可能性は高い。今後、図書館や各種アーカイブが受け入れた電子資料の内容を長期的に保存するためには、プロテクトの解除やライセンス上の問題の解決など、資料受入の段階においてマイグレーションの妨げにならないようにする対策を取ることが必要であると考えられる。

5.2 エミュレーションによるアクセス手段の確保

4.2のエミュレーション調査では、エミュレーションが、現在の技術水準では問題の根本的な解決に結びつかないことが明らかになった。特に音声と画像の問題が大きく、全体として完全に再生できた資料の割合は3割に留まった。これは調査で使用したエミュレータ固有の問題である可能性があるが、音声などを完全に再現することの難しさを表しているといえる。現時点では、完璧なエミュレーションは困難だと考えざるを得ない。また、市販のエミュレータは、より需要の高い環境を再現するように開発される傾向があり、同じエミュレータでもバージョンが上がると旧式 OS 用のエミュレーション能力が下がる可能性がある。今回の再生結果でも、より古い世代の OS のほうが再生できた割合が少ないことに、その傾向が現れている。

今回の調査の結果を踏まえ、電子資料への長期的アクセスを確保するためには、エミュレータは、以下の要件を備えていることが必要であると考えられる。

- ハードウェア環境や OS 環境に依存せずに旧式ハードウェア環境を再現できること
- 各 OS に対応した各種デバイスドライバが入手可能で、旧式ハードウェア環境を再現できること
- エミュレータがバージョンアップされても、旧バージョンがサポートしていた旧式 OS 環境を引き継いでサポートしていること

しかし、通常はエミュレータ自体も特定の OS で動作するために、そのままでは長期的に使い続けることはできない。このため、エミュレータ上で旧式のエミュレータを動作させることや、多くのハードウェア上で比較的容易に実装可能な仮想マシン上で動作可能なエミュレータを開発することなどが必要と思われる。

5.3 ファイル形式変換

4.3のファイル形式変換調査では、データのみ電子資料のサンプル100点について調査したところ、ファイル形式変換の対象となった資料は15点と予想外に少なかった。その中で変換および再生確認ができた資料は2点に留まった。対象外となった電子資料の大多数が、HTMLなどの複数のファイルが関連性を持ちながらコンテンツとして再生される資料であり、これらに対しては、ファイル単位の変換ではなく電子資料単位で、複数ファイルの関連性が維持されるように変換を行う必要がある。この場合、例えば始点となるファイルを指定するなど、電子資料単位での判断が必要となる。しかし、現在のファイル形式変換等の技術では困難であり、長期保存の技術としては補助的な手段に過ぎないといえる。現状では、アプリケーションプログラムを入手してデータと共に保存することが、有用な手段であると考えられる。そのためには、資料とアプリケーションプログラムを一元的に管理する仕組みが必要である。

6. 今後の取り組み

本調査を通じて、電子情報の長期的保存とアクセス手段の確保には様々な問題があることが明らかになった。今後、具体的な取り組みを進めるには、次に挙げる3点の課題を解決することが必要である。

(1) 長期保存と利用のためのシステム構築

長期保存を行うためには、それを可能とするためのシステムを構築しなくてはならない。デジタル情報の長期保存のための指針として OAIS 参照モデル [8] が提案されているが、長期保存と長期的な利用を可能とするシステムとして具体化することは容易ではない。また、OS のバージョンアップの頻度、技術進歩にシステム全体を適合させることを考慮すれば、4～5 年毎に全面的なシステムの更新を行いつづけてはならない。このような非常に移ろいやすい基盤の上に、「永続的な長期保存システム」を構築せざるを得ないため、システムの構築以上に、システムが備えるべき永続的な特質についての検討や、システムを状況に適応させ続けていくための戦略が重要である。

(2) 技術動向の把握

コンピュータとその関連技術が進歩するに伴い、NDL が受け入れる電子出版物も多種多様になっている。ハードウェア・ソフトウェア・通信・出版技術等、常に最新の情報を基にして、新たな種類の電子媒体に対しても確実に対応する必要がある。「気付かぬうちに技術的環境に変化が起こり、大量の電子資料が読み取れなくなってしまった」という事態は避けなければならない。

(3) 協力関係の確立

コピープロテクトの解除、エミュレータの作成等、図書館が単独で行うには困難な作業が多く、パッケージ系電子出版物の著作者や技術提供者に問題を周知し、適切な協力を得る必要がある。また、国内および世界各国で同様の研究を行っている国立図書館との連携も重要である。

NDL は、2004 年 2 月に「電子図書館中期計画 2004」を策定し、目標の一つとしてデジタル・アーカイブの構築を挙げている [9]。電子図書館事業の基盤技術の一つとして、今後も調査研究を進めると共に、具体的な長期保存の取り組みのための枠組みを構築していきたい。

(注) パッケージ系出版物は、下記の 3 種に分類される。

- 電子資料：映像資料、録音資料以外の資料。PC やゲーム専用機などで再生されることを前提に作成された資料。
- 映像資料：ビデオテープ、ビデオディスク、DVD など映像を記録した資料。
- 録音資料：CD、DVD-Audio、MD、レコードなど音声（主に音楽）を記録した資料。

このうち、電子資料について調査を行った。

参考文献

- [1] UNESCO Charter on the Preservation of the Digital Heritage http://portal.unesco.org/ci/en/ev.php-URL_ID=13366&URL_DO=DO_TOPIC&URL_SECTION=201.html [last access: 2006.2.13]
- [2] 国立国会図書館: 納本制度とは http://www.ndl.go.jp/jp/aboutus/deposit_01what.html [last access: 2006.2.13]
- [3] 国立国会図書館: 近代デジタルライブラリー <http://kindai.ndl.go.jp/> [last access: 2006.2.13]
- [4] 国立国会図書館: 貴重書画像データベース <http://rarebook.ndl.go.jp/> [last access: 2006.2.13]
- [5] 国立国会図書館: インターネット資源蓄積実験事業 (WARP) <http://warp.ndl.go.jp/> [last access: 2006.2.13]
- [6] 国立国会図書館: 電子情報の長期的な保存と利用 <http://www.ndl.go.jp/jp/aboutus/preservation.html> [last access: 2006.2.13]

[7] これまでに様々な調査機関が結果を提示している。例として、

Fred R. Byers: "Care and Handling of CDs and DVDs -A Guide for Librarians and Archivists", NIST Special Publication 500-252, National Institute of Standards and Technology and Council on Library and Information Resources, 2003.

<http://www.itl.nist.gov/div895/carefordisc/CDandDVDCareandHandlingGuide.pdf>

C. Shahani: "Compact disc service life: an investigation of the estimated SERVICE life of prerecorded compact discs (CD-ROM)", Preservation Research and Testing Series, no. 10, Preservation Directorate Library of Congress, 2005.

[8] "Reference Model for an Open Archival Information System (OAIS). Blue Book, Issue 1.", CCSDS 650.0-B-1, Consultative Committee for Space Data Systems, 2002.

<http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf> [last access: 2006.2.13]

[9] 国立国会図書館:電子図書館中期計画 2004 http://www.ndl.go.jp/jp/aboutus/elib_plan2004.html [last access: 2006.2.13]